DOCUMENT RESUME

ED 312 304 TM 014 092

AUTHOR Burry, Judith A.; Shaw, Dale
TITLE Defining Teacher Effectiveness on a Continuum: A Rasch Model Approach.
PUB DATE Apr 88
NOTE 32p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 6-8, 1988).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Elementary School Teachers; Elementary Secondary Education; Latent Trait Theory; Public Schools; *Rating Scales; *Secondary School Teachers; Teacher Attitudes; *Teacher Effectiveness; Teacher Evaluation; Teacher Student Relationship; *Test Format
IDENTIFIERS *Inventory of Myself as a Professional; Observation Techniques; *Rasch Model; Self Report Measures

ABSTRACT

The Rasch rating scale model was used to provide diagnostic information about a teacher effectiveness instrument, the Inventory of Myself as a Professional (IMP). Two forms of the instrument, a self-report form to be completed by the teacher and an observation form to be completed by a supervisor or administrator contained the same thirty-six items. The items were included on three subscales: (1) student to teacher relationships; (2) professional attitude; and (3) professional skills. Response patterns were scored on a scale from one to five. The effectiveness of 204 teachers as observed by their principals on the Observation Form of the IMP and the effectiveness of 211 teachers as reported on the Self-Report Form of the IMP were assessed. Data for both forms were collected for teachers from public elementary and secondary schools in Colorado. Both forms appeared to fit the Rasch model quite well. The teacher effectiveness variable was defined on both instruments, providing construct validity to the IMP. Item calibrations sufficiently spread the person measures to ensure an adequate fit of the model on both forms. Having person measures contributed to the diagnostic potential of the IMP. Results suggest that the Rasch model provides a major breakthrough in analyzing the variable of teacher effectiveness in the use of item calibrations on a hierarchical continuum. Instructions for both forms of the instrument are included. A table and six figures illustrate the study results. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED312304

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Judith A. Burry

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

DEFINING TEACHER EFFECTIVENESS ON A CONTINUUM:

A RASCH MODEL APPROACH

Judith A. Burry

University of Alabama

Dale Shaw

University of Northern Colorado

Paper presented at the National Council on Measurement in Education
New Orleans, April, 1988

TM014092

BEST COPY AVAILABLE

SUMMARY

The Rasch Rating Scale Model is applied to provide diagnostic information on a teacher effectiveness instrument called the Inventory of Myself as a Professional (IMP). Thirty-six items are included in three subscales: student teacher relationships, professional attitude, and professional skills. Items are identical on two separate forms of the instrument: a Self-Report Form, and an Observation Form. Response patterns are scored on a scale of one to five.

Data for both instruments were collected on 211 teachers whose school buildings were randomly selected from the state of Colorado public schools, grades K-12, as part of a validation study conducted at the University of Northern Colorado and endorsed by the Colorado Department of Education.

Item and person analyses are positioned on an attitudinal scale in logits. Scales are analyzed separately on each form by the items positions at their calibrations on the attitude variable. Persons are ordered and compared by professional level elementary and secondary.

Items and persons are compared within tests and across both forms. The item difficulties and person measures provide insight into the dimension of teacher effectiveness not previously available in teacher assessment. Diagnostic implications of both parameters allows for new application of the teacher effectiveness dimension.

INTRODUCTION

Teacher effectiveness is currently a prominent national issue. Soar, Medley, and Coker (1983) claim the need for better ways of evaluating teachers to identify qualified teaching personnel as urgent. Bennett (1978) suggests that the current methods of teacher evaluation are ineffective. Current research methods have not provided specific direction. This lack of specificity is due partly to the nature of the definition of teacher effectiveness and a lack of consensus in the research findings.

Borich (1977) reported no scaled diagnostic measures in teacher effectiveness in his review of teacher evaluation instruments from the Ryans study in the 1950s through the middle 1970s. Several states have responded to mandated legislation for teacher certification with their own performanced-based instruments. Burry (1984) reviewed the psychometric properties of these instruments and found that the reliability coefficients did not exceed .77 for a total instruments, and the scales of these instruments to be considerably lower. Validity studies were limited to content validity. Attempts were made to establish construct validity through factor analysis; however, none of the scales were supported by this statistical procedure. There appears to be a lack of consistency in what the experts in the field judge to be one dimension and what the statistical

procedures define as another. These congruencies only contribute to the lack of consistency and specificity needed to provide a clear, practical, and workable definition of teacher effectiveness.

Perhaps part of the problem lies in the instruments themselves and the other part of the models applied in the validation process. At this time the literature does not report an effective use of the Rasch Model in validating the domain of teacher effectiveness. Use of the present instrument results is limited to pass-fail, certified-not certified, hire-fire, or more-no additional monies concepts rather than how can the teacher(s) develop into a more effective professional. Wright and Masters (1982, p.90), state, "The purpose of a measurement model is to extract from suitable data a useful definition of an intended variable and then to measure persons on this variable." The Rasch Model attempts to verify that the items on an instrument work together to define the variable. Applying the Rasch Model to a teacher effectiveness instrument would provide not only the validation of the teacher effectiveness dimension, but item calibrations and person ability measures. This information would be most useful in providing diagnostic information.

INSTRUMENT DEVELOPMENT

The IMP was created to attempt to provide substance and clarity to the definition and dimension of teacher effectiveness

along with diagnostic information. All of the items and the scales were constructed based upon the literature. The three scales are: 1) Teacher-Student Relationships, 2) Professional Attitude, and 3) Teaching Skills. The two forms were developed (self-report and an observation instrument) to be used by a supervisor or an administrator. These forms are identical with the exception of the preface statement. The self-report instrument asks the teacher to rate themselves on items prefaced by, "AS A CLASSROOM TEACHER," while the Observation Form asks the rater to respond to the items prefaced by, "THIS CLASSROOM TEACHER."

All 36 items are of a uniform style and were constructed on a Summated Rating Scale. Scale responses are: (A) almost never true, (B) seldom true, (C) true about half of the time, (D) true most of the time, and (E) almost always true. Responses were converted to a numerical scale where (A) almost never true has a weight of 1, (B) seldom true has a weight of 2, on up to (E) almost always true which has a weight of 5. Item responses were summed into scales and a total instrument value.

After a number of item trials and instrument revisions, a final validation study was done using a systematic, stratified sample of 150 schools systematically and randomly selected in the state of Colorado. There were 204 teachers who were observed by their building principal and 211 teachers who responded to the self-report instruments.

RASCH RATING SCALE MODEL

Because the response format of the IMP is a Summative Rating Scale, Andrich (1978a, 1978b) suggests that an appropriate model for the ordered response categories is the Rasch Rating Scale Model. He comments that in order to overcome the objections to the integer-scoring procedures, a response model which keeps track of the person's categorical responses is needed. Andrich (1978b, p. 581) states, "a multidimensional parametric structure is postulated with respect to the response categories and after the parameter estimation is carried out, checks on the possible reduction of the dimensionality of this structure are made."

According to Wright and Master (1981), the relative difficulties of the "steps" on a rating scale are to be governed by the fixed rating points accompanying the items. Assuming the response format is the same, these steps should be invariant. The pattern of the steps at a location is determined by the item's response set and is estimated once for the complete set of items. Items then differ only by their location on the variable of interest.

The rating scale is identified by Andrich (1978a, 1978b) and can be written as the probability of a person n responding in category x to item i.

$x=0,1,\ldots,m$ Rating Scale

$$\pi_{nix} = \frac{\exp \sum_{j=0}^{x} [\beta_n-(\delta_i+\tau_j)]}{\sum_{k=0}^{m} \exp \sum_{j=0}^{k} [\beta_n-(\delta'+\tau_j)]}$$

where $\tau_0=0$ so that $\exp \sum_{j=0}^{0} [\beta_n-(\delta_i+\tau_j)] = 1.$

When a rating scale is analyzed using this model, Wright and Master (1981) say that it is possible to obtain an estimate for each person n, and estimated scale value for each item and estimates of the response "thresholds" $\tau_1, \tau_2, \ldots \tau_m$ for m + 1 rating categories. Because these kinds of information are invaluable for utilizing teacher effectiveness characteristics for diagnostic purposes, the Rasch Rating Scale Model was applied to the IMP data on both forms of the instrument.

IMP DATA ANALYSIS

To determine whether a measurement model fits the data, it is necessary to find out whether or not the items in the instrument are working together, to define the variable, and to what extent the person's responses fit the expectations of the model. The degree of precision to which the hierarchical arrangement of the item calibrations expressed in logits on the dimension supports the teacher effectiveness variable provides evidence for the construct validity of the IMP. Another important aspect of the data analysis is to determine the validity and placement of each person's measure and to observe whether or not these persons are separated along the dimension of teacher effectiveness. The extent to which the item calibrations and persons statistically fit the model is, in turn, representative of

the ability of the IMP to provide a diagnostic index of teacher effectiveness items to spread the teachers along the dimension. Therefore, the results of the IMP data analysis using the Rasch Scaling Model will provide varying degrees of instrument, item, person, and variable validity.

Verifying the Variable Teacher Effectiveness of the IMP

The effectiveness of 204 teachers as observed by their principals on the Observation Form of the IMP and the effectiveness of 211 teachers as reported on the Self-Report Form of the IMP both on the same 36 items, is discussed in this section of the paper. Each form of the instrument will be discussed first separately and then compared with each other on item calibrations positions to define the variable teacher effectiveness.

Table 1 presents the scale item statistics in sequence order for both instruments. The Item names ST, PA, and TS refer to the three scales of the instruments: student teacher relationships, professional attitude, and teaching skills. Again, the scales are the same for both forms of the instrument. The items are presented at the end of this paper. It is interesting to note that there is a smaller standard deviation of the items and the standard deviation of the fit statistic on the Observation Form. The mean for the error is the same.

Figure 1 lays out the teacher effectiveness variable on the Observation Form. The items are presented in logits along the

dimension for the total instrument as well as each of the three scales. The items are positioned at their calibrations and the semicircles represent the standard error of each item along the continuum. There is little deviation in the standard error of the items. See Table 1. The items positioned at the top are the easier characteristics (variables) in teacher effectiveness on which to observe teachers. The entire continuum is represented between -1.0 and +1.5. Six of these items have fit statistics greater than $\pm$4.0 which is of some concern.

Note that many of the easier items are on the first scale which is student teacher relationships. Item 17 on the Observation Form is, "Enjoys working with students in the classroom." Item 30 is, "Enjoys interacting with students." Item 23 is, "Provides opportunities for pupils to experience success." Item 23 is, "Provides opportunities for pupils to experience success." All of these item calibrations position themselves between -1.0 and the mean 0 and they cluster around the same f calibration points. Scale 1 appears to be an easier scale on which to observe teachers. It has an item calibration mean of -.399.

See Table 2.

TABLE 2

LOGIT MEANS AND STANDARD DEVIATIONS

FOR THE IMP SCALES

Self-Report Form

Observation Form

	Scale 1	Scale 2	Scale 3	Scale 1	Scale 2	Scale 3
mean	-.674	.093	.364	-.339	.092	.153
SD	.459	.974	.495	.247	.577	.403

Scale 2 is a Professional Attitude. For the most part, these items represent the teacher's attitude toward his/her profession. Item 34 on the Observation Form reads, "Am committed to the primary goal of student growth." Item 24 is, "Exhibits businesslike or task-oriented behavior." The item calibration positions are along the entire instrument continuum and the mean of the item calibrations on these items is .092. This scale appears to be representative of the whole scale and is slightly more difficult for administrators to evaluate teachers.

The third scale is Teaching Skills and it has an item calibrations mean of .153 which is the most difficult of the three scales on which administrators evaluate teachers. These 16 items also cluster around many of the same calibration points. Item 3 on Scale 3 is, "Plans interesting lessons." Item 5 is, "Relates

instruction to instructional objectives."

Items on the Self-Report Form cover more of the teacher effectiveness variable. It is interesting to note that the same items calibrate in a similar pattern on both instruments, but are spread out more on the Self-Report Form. The entire range is represented on the variable from about -1.6 to 2.0. There is a large break in the continuum between items 25, 4, and 12 which is a concern. The standard error is greater on these easier items. There are only two items which fit statistics greater than ± 4.0 and would be candidates for revision or deletion.

The first scale on the Self-Report Form has an item calibration mean of -.674, a wider range and more breaks in the continuum than does 4 on the Observation Form. Again this scale appears to be the easiest scale on which teachers evaluate themselves.

Scale 2 is more spread out along the dimension than either of the other two scales. It also has more breaks in the continuum than it does on the Observation Form. The mean of the item calibrations is .093 which is almost identical to the mean on the Observation Form and is slightly more difficult than the first scale.

Scale 3 lays out over the more difficult end of the continuum with a calibrated logit mean of .364. The item calibrations position themselves without a break in the continuum.

Comparing the scales across forms, the item calibrations are similar but not identical, and none of the items change positions drastically. There are several items with unduly large fit statistics and are candidates for deletion or revision from their respective instruments.

Item calibrations on both instruments are spread along the dimension more so on the Self-Report Form than on the Observation Form. There is only a slight break in the continuum on the Observation Form and one large break on the Self-Report Form. Ideally there would be no breaks and both instruments would have items that have calibrated positions between ± 2.00. Nonetheless, both instruments have item calibrated positions that sufficiently cover the dimension of teacher effectiveness. The scales position themselves in a similar pattern along the dimension even though the variance is greater on the Observation Form. The item calibration positions on these instruments suggest evidence for construct validity. Figure 3 is a graphic representation of the comparison of the item difficulties of the two instruments. The correlation coefficient is .74 with a slope of .44 and an intercept of .03. The relationship is linear which also supports the construct validity of the dimension teacher effectiveness in both instruments.

Identifying Person Measures Along the Teacher Effectiveness Variable

Identifying Person Measures Along the Teacher Effectiveness Variable

Person measures are spread along the logit scale from 6.67 to -3.85 on the Observation Instrument in Figure 4 (see columns 1 and 2). Column one lists the person's raw score, the number of persons with that raw score, the person's measure position, and the error for that particular measure. Persons more toward the top of the scale are observed to be more effective teachers by their principal, whereas persons at the bottom of the scale are individuals perceived by the principal as having less teacher effectiveness ability. It is interesting to note the distribution of the persons on the variable in center column. The people have a mean logit of 1.72 with a standard deviation of 2.07 along the 36 items. They are spread out to form a slightly negative platekurtic distribution. However, the persons are spread out along the variable, which is essential for the model to fit the data.

There are nine persons with fit statistics greater than 4.0. These people are candidates for deletion on the next data analysis. Person errors are not exceptionally high except on the high extreme of the distribution. The reliability of person separation is .98. This means that the Rasch Scaling Model is suitable for these items on observed teachers using the IMP Observation Form. Interesting to note, the item calibration positions in the last column compared to the person measure

positions in the middle column. The item calibrations cluster themselves more in the center and to the less able, easier, end of the variable. There is much more variability among the persons than the items.

Person measures on the Self-Report Form have an entirely different distribution (see Figure 5). These measures position themselves with a mean of 1.93 and a standard deviation of .87. The curve appears skewed positively, but is much more normal than the Observation Form. However, there is much less variability among the subjects. This is understandable since self-report data has more of a tendency to be less variable with higher scores than observation data of similar content. There are four persons with fit statistics greater than 4.0. Again these people would be candidates for deletion on the next analysis. The reliability for the person separation on the Self-Report Form is .90. The Rasch Scaling Model is also suitable for these 35 items on the Self-Report Form of the IMP.

The item calibration positions are spread out more on the Self-Report Form than they are on the Observation Form, while the persons have less variability on the Self-Report Form than they do on the Observation Form. These data support the objectivity of the model to separate persons and items independently of each other.

Teachers Across Elementary and Secondary Levels

Note the boxes around the person frequency distribution on Tables 4 and 5. The boxed person is an elementary teacher responsible for students in grades kindergarten through grades 5 or 6, depending on the configuration of the school building in which the person teachers. The unboxed persons are secondary teachers. Persons in both levels appear to be evenly spread out along the continuum on the Observation Form. The Self-Report Form has more elementary teachers at the high end of the variable and fewer at the low end of the variable. There appears to be more of a discrepancy between professional levels on the Self-Report Form than on the Observation Form. According to Burry (1984) the items were originally written to be generic for both elementary and secondary teachers. Further analysis of persons and items is needed to conclude that indeed these items are generic in that they are interpreted the same at both professional levels.

Probability Curves for Five Response Categories on the IMP

The category probability curves for both forms of the IMP are presented in Table 6. Responses for both instruments were (A) almost never true, (B) seldom true, (C) true about half the time, (D) true most of the time, and (E) almost always true. The curves with the 0's represent the almost never true; the 1's the seldom true; the 2's true about half the time; the 3's true most of the time; and the 4's almost always true. The numbers are down-shifted one from their raw score assignment in the computer

program for the Rasch Rating Scale Analysis. The X axis represents the person's ability in logits and the Y axis is the probability of the person with a specific ability responding to each of the successive categories. For example, a person with a measure of -2.0 on the Observation Form would have the probability of being rated seldom true 50% of the time. A person with a measure on each instrument at 0.0 would have a probability of 57% of being rated true about half the time on the Observation Form and a probability of 50% of rating themselves true about half the time on the Self-Report Form.

The probability curves are very different for each instrument which endorses the fact that although the items are the same, the perceptions of these items in their application are very different; there are two separate instruments.

SUMMARY AND CONCLUSIONS

The major question of investigation was whether the Rasch Rating Scale Model fit one or both of the forms of the teacher effectiveness instrument The Inventory of Myself as a Professional (IMP). Because the data in the literature in its current state of the art do not utilize the use of the Rasch Rating Scale Model, it is essential to answer this question first. It seems that both forms of the instrument appear to fit the model quite well. For the most part, item calibrations on both instruments had a minimum of standard error and only a few items with exceptionally high fit

statistics. The items on both instruments sufficiently spread the persons along the variable teacher effectiveness. Only a few persons had extremely high fit statistics above 4,000. The reliability coefficient of person separation is .98 on the Observation Form and .90 on the Self-Report Form. It appears that the Rasch Rating Scale Model is most appropriate for analyzing the IMP on both instruments.

Other questions of investigation ask to what degree is the variable teacher effectiveness validated by the use of the model? Do the item calibrations and the person measures work together to define the variable on both forms of the instrument? Does the hierarchical arrangement of the item calibrations on the teacher effectiveness variable separate the persons along the same variable?

The variable teacher effectiveness is defined in a small range from -.72 to 1.26 on the Observation Form by the hierarchical positions of the item calibrations. There is only 1 small break at 1.0 on the continuum. The teacher effectiveness variable is defined. However, the definition could be strengthened by adding items with calibrated positions at both ends of the variable. The range on the Self-Report Form is wider from -1.32 to 1.91. Again, the variable is defined; however it could be strengthened by adding items with calibrated positions at this spread and the easier end of the continuum. The teacher-effectiveness variable is defined on both instruments and thus

providing construct validity to the IMP.

The item calibration positions spread out along the variable with small standard errors of measurement. Both the Self-Report and the Observation Forms of the IMP provide a definition of the teacher effectiveness variable. The scales on both instruments position themselves in the same order of degree of difficulty on the item calibrations. This is of significance to teacher education programs in that the student teacher relationship scale is the easiest, the professional attitude scale is in the middle position, and for the most part, the most difficult items are the teacher skills. The teaching skills can and are taught in most teacher education programs. However, until this analysis was complete there has been no confirmation that these skills were any harder to perform or self-analyze than any other kind of skills in teacher effectiveness. This is an intriguing discovery.

The item calibrations on both instruments spread out the person measures with a minimum of error and misfitted the persons. The distributions are different; the Observation Form has a much greater variance on the person measures than does the self-report instrument. However, both populations are adequately spread out by the item calibrations to obtain adequate person measures. As Wright and Masters (1981) comment, with the use of the Rasch Rating Scale Model it is possible to obtain interval measurements of persons' responses to a Summative Rating Scale. This is a major contribution and improvement to the field of teacher

evaluation.

The analysis of the category probability plots supports the use of each instrument as a separate instrument, as the response patterns are very different. Burry (1984) concludes that little has been done to validate the use of the same items with two different applications.

To summarize the results the Rasch Rating Scale Model fits both the observation and Self-Report Form of the IMP. The item calibrations sufficiently spread the person measures to ensure an adequate fit of the model on both forms. The variable teacher effectiveness is defined with some limitations.

Further investigations are needed to analyze the data across teaching levels. Item calibrations for each level would provide more insight into how teachers at each level view each item of teacher effectiveness and whether or not the dimensions of the scales change or remain constant.

In all, the Rasch Rating Scale Model provides a major breakthrough in analyzing the variable of teacher effectiveness in the use of items calibrations on a hierarchical continuum. Having person measures contributes to the diagnostic potential of the IMP and other similar scales. Instruments with good psychometric properties can only facilitate and enhance the potential and application of the teacher effectiveness variable in training and assisting the development of effective professional teachers.

REFERENCES

Andrich, D. (1978a). Scaling attitude items constructed and scored in the Likert tradition. Educational and Psychological Measurement, 38, 665-680.

Andrich, D. (1978b). Application of a psychometric rating model to ordered categories which are scored with successive integers. Applied Psychological Measurement, 2, 581-594.

Bennett, N. (June, 1978). Recent research on teaching: A dream, a belief, and a model. British Journal of Education Psychology, 5-37.

Borich, G. (1977a). The appraisal of teaching: Concepts and process. Addison-Wesley Publishing Company.

Burry, J. (1984). Teacher effectiveness in development and validations of the inventory of myself as a professional (IMP). Unpublished doctoral dissertation. The University of Northern Colorado.

Soar, R., Modley, D., & Coker, H. (1983). Teacher evaluation: A critique of currently used methods. Phi Delta Kappan, 65(4).

Wright, B. & Mastons, G. (1981). Rating scale analysis rasch measurement, MESA Press, Chicago, Illinois.

INSTRUMENTS

ASSESSMENT BY AN OBSERVER

INVENTORY OF MYSELF AS A PROFESSIONAL (IMP-A)

*****

The Inventory of Myself as a Professional (IMP) was developed to obtain information about the classroom teacher's professional performance. There are no right or wrong answers. The purpose of the IMP is to provide feedback for the individual teacher's professional growth.

Each of the statements asks you to describe the professional classroom teacher and is prefaced with THIS CLASSROOM TEACHER IS . . . . On the other side of the form are statements which ask you to describe the classroom teacher that you are observing. Respond to each statement with these letters A, B, C, D, or E.

A. ALMOST NEVER TRUE The statement about what the professional classroom teacher says or does is almost never true.
B. SELDOM TRUE This statement about what the professional classroom teacher says or does is true only occasionally.
C. TRUE ABOUT HALF OF THE TIME The statement about what the professional classroom teacher says or does is true about half of the time.
D. TRUE MOST OF THE TIME The statement about what the professional classroom teacher says or does is almost never true.
E. ALMOST ALWAYS TRUE The statement about what the professional classroom teacher says or does is true almost always.

DIRECTIONS: Read each statement quickly and carefully. Select the letter which best describes the classroom teacher. The letters are on the left side of each statement. Completely blacken the letter which corresponds to your answer. Do not just circle the letter. Blacken only 1 letter for each statemtnt.

PLEASE USE A NUMBER 2 PENCIL

*****

Prepared by: Judy Burry, Dale Shaw

Copyright: 1983

SELF ASSESSMENT

INVENTORY OF MYSELF AS A PROFESSIONAL (IMP-SA)

*****

The Inventory of my Myself as a Professional (IMP) was developed to obtain information regarding how the classroom teacher feels about his/her professional performance. It is self-administered; there are no right or wrong answers. The purpose of the IMP is to provide feedback for your individual professional development.

Each of the statements asks you to describe yourself as a classroom teacher and is prefaced with "AS A CLASSROOM TEACHER, I . . ." Respond to each statement with one of these letters A, B, C, D, or E. On the other side of the form are statements which ask you to describe yourself.

A. ALMOST NEVER TRUE The statement about what I say or do as a classroom teacher is almost never true.
B. SELDOM TRUE The statement about what I say or do as a classroom teacher is true only occasionally.
C. TRUE ABOUT HALF OF THE TIME The statement about what I do as a classroom teacher is true about half of the time.
D. TRUE MOST OF THE TIME The statement about what I say or do as a classroom teacher is true most of the time.
E. ALMOST ALWAYS TRUE The statement about what I say or do as a classroom teacher is almost always true.

DIRECTIONS: Read each statement quickly and carefully. Select the letter which best describes you as a classroom teacher. Completely blacken the letter which corresponds to your answer. Do not just circle the letter. Blacken only 1 letter for each statement.

PLEASE USE A NUMBER 2 PENCIL

*****

Authors: Judy Burry, Dale Shaw

Copyright: 1983

TABLE 1

SCALE ITEM STATISTICS
SERIAL ORDER

SELF REPORT FORM

SERIAL ORDER

SEQ NUM	NAME	ITEM VALUE	ERROR	SFIT
1	STO1	-0.33	0.11	-2.24*
2	PAO1	0.70	0.09	-2.10*
3	TSO1	0.66	0.09	-4.98*
4	TSO2	1.07	0.08	-0.98
5	TSO3	-0.28	0.11	-1.21
6	STO2	-0.46	0.12	0.26
7	TSO4	0.01	0.11	3.31*
8	STO3	-0.04	0.11	-1.05
9	PAO2	0.90	0.09	4.65*
10	TSO5	-0.55	0.12	-1.03
11	TSO6	0.92	0.09	-0.20
12	PAO3	1.91	0.08	4.55*
13	TSO7	-0.23	0.11	-4.01*
14	STO3	-0.96	0.14	-1.49
15	PAO4	-0.77	0.13	-2.62*
16	TSO9	0.23	0.10	-5.17*
17	STO5	-1.15	0.14	-2.26*
18	PAO5	0.59	0.09	3.22*
19	TSO9	0.47	0.10	2.75*
20	PAO6	-0.63	0.12	-2.61*
21	TS10	0.25	0.10	-1.05
22	TS11	0.85	0.09	7.77*
23	STO6	-0.58	0.12	-3.82*
24	PAO7	-0.05	0.11	-0.16
25	TS12	1.07	0.08	0.37
26	TS13	0.78	0.09	-3.08*
27	STO7	-0.40	0.12	-4.20*
28	PAO8	-0.59	0.12	-1.15
29	STO8	-1.26	0.15	-0.56
30	STO9	-1.32	0.15	-0.58
31	PAO9	0.26	0.10	-5.61*
32	TS14	0.29	0.10	0.14
33	STO7	-0.23	0.11	-3.03*
34	PA10	-1.39	0.16	-1.69
35	STO9TS	0.10	0.10	2.67*
36	TS16	0.19	0.10	0.28

OBSERVATION FORM

SERIAL ORDER

SEQ NUM	NAME	ITEM VALUE	ERROR	SFIT
1	STO1	-0.19	0.11	0.60
2	PAO1	0.47	0.11	-1.20
3	TSO1	0.06	0.11	-3.89*
4	TSO2	1.13	0.10	1.84
5	TSO3	0.02	0.11	-0.99
6	STO2	-0.37	0.11	-2.62*
7	TSO4	-0.20	0.11	2.62*
8	STO3	0.04	0.11	-0.93
9	PAO2	0.66	0.11	0.94
10	TSO5	-0.25	0.11	0.90
11	TSO6	0.53	0.11	2.28*
12	PAO3	1.26	0.10	5.16*
13	TSO7	-0.47	0.11	-2.70*
14	STO3	-0.38	0.11	-2.01*
15	PAO4	-0.30	0.11	-2.80*
16	TSO9	-0.02	0.11	-1.59
17	STO5	-0.72	0.12	-3.24*
18	PAO5	0.08	0.11	3.64*
19	TSO9	-0.10	0.11	0.62
20	PAO6	-0.34	0.11	-0.60
21	TS10	0.18	0.11	-1.88
22	TS11	0.21	0.11	3.12*
23	STO6	-0.60	0.12	-3.09*
24	PAO7	-0.32	0.11	2.07*
25	TS12	0.42	0.11	0.86
26	TS13	0.77	0.11	-2.03*
27	STO7	-0.17	0.11	-4.34*
28	PAO8	-0.10	0.11	0.77
29	STO8	-0.18	0.11	-2.82*
30	STO9	-0.64	0.12	-0.68
31	PAO9	0.20	0.11	-3.55*
32	TS14	0.23	0.11	-1.13
33	STO7	-0.18	0.11	-3.04*
34	PA10	-0.69	0.12	-2.94*
35	STO9TS	-0.12	0.11	-0.51
36	TS16	0.06	0.11	-3.30*

SUMMARY STATISTICS FOR SCALE ANALYSIS

	VALUE	ERROR	SFIT
MEAN	-0.00	0.11	-0.75
SD	0.77	0.01	2.98

	VALUE	ERROR	SFIT
MEAN	-0.00	0.11	-0.73
SD	0.47	0.0	2.38

FIGURE 1

DEFINING THE IMP VARIABLE BY ITEM DIFFICULTY OBSERVATION FORM

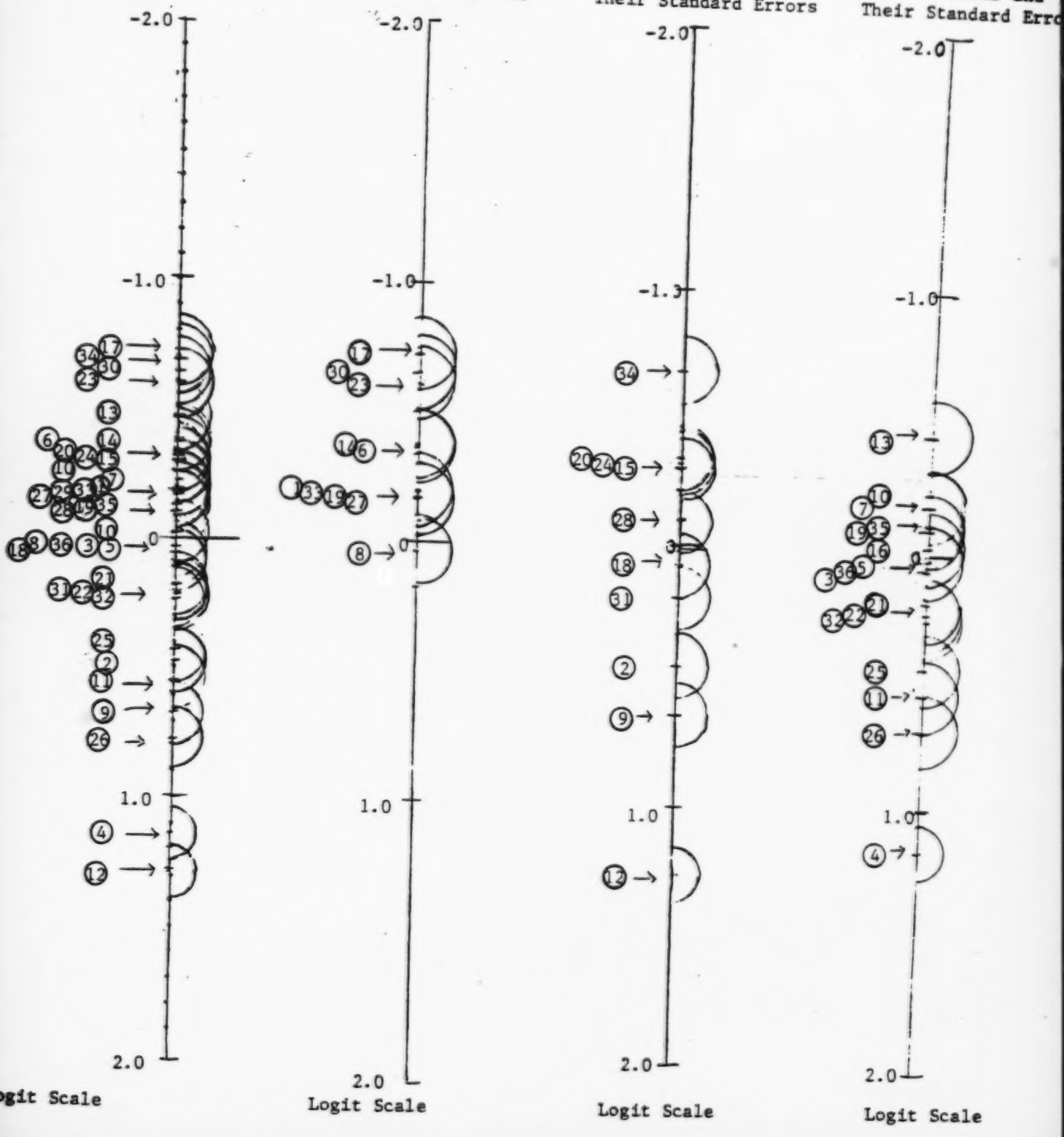


FIGURE 2

DEFINING THE IMP VARIABLE BY ITEM DIFFICULTY
SELF-REPORT FORM

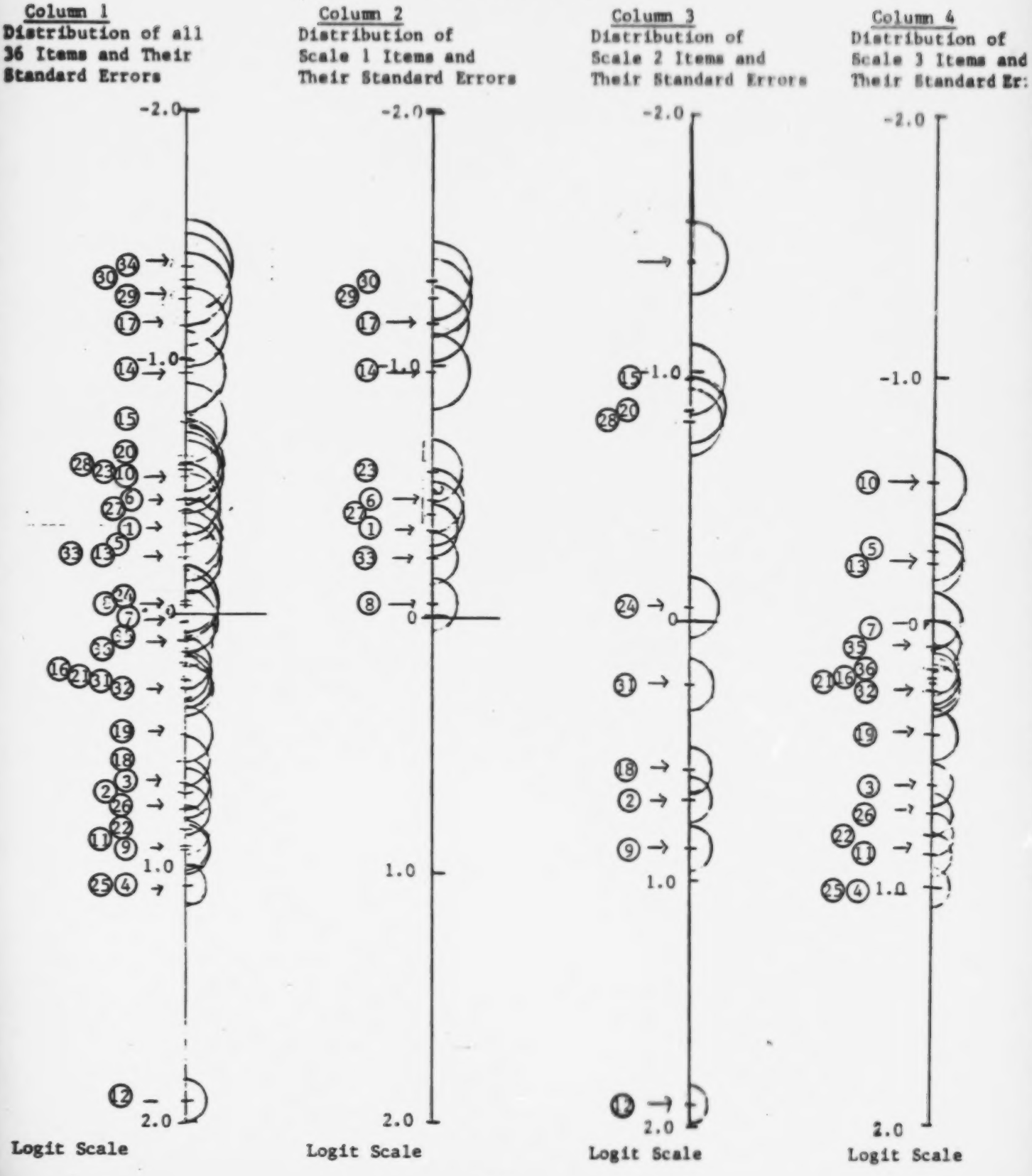


FIGURE 3

COMPARING ITEM DIFFICULTIES OF
THE IMP OBSERVATION AND SELF-REPORT
FORMS

36 ITEMS

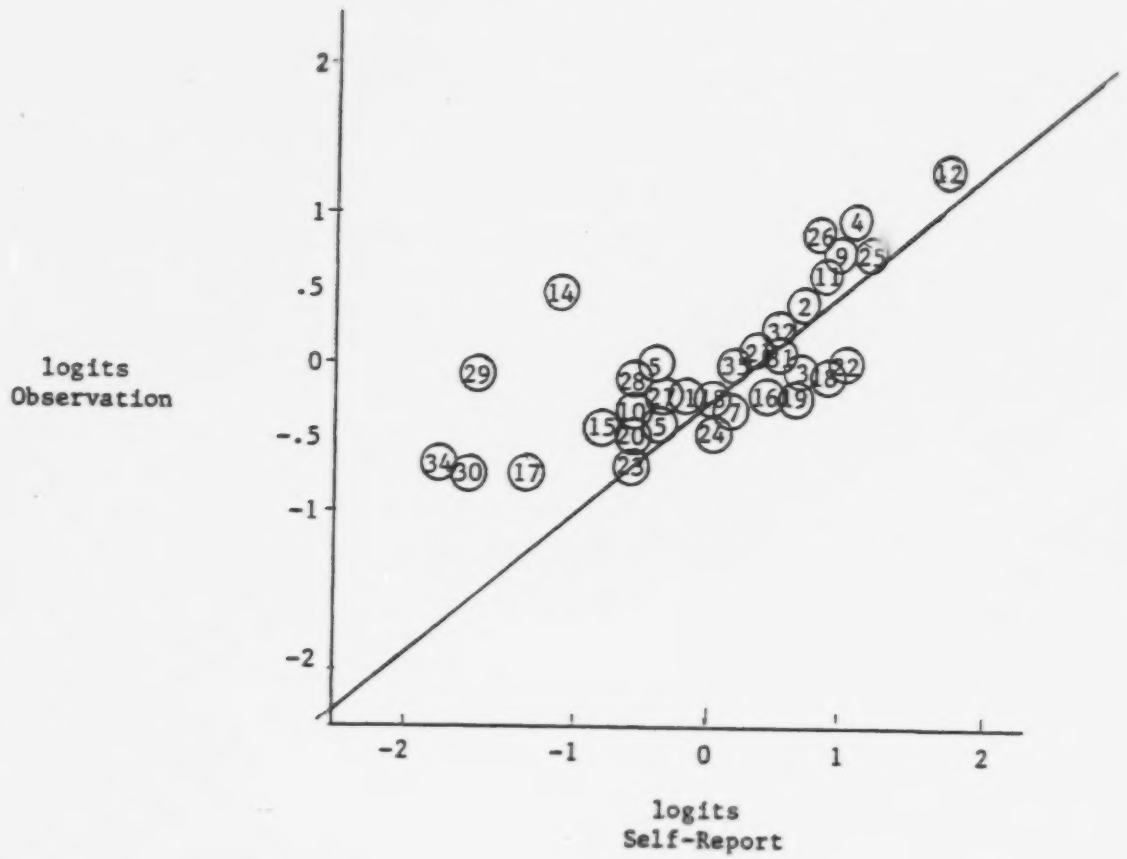


FIGURE 4
SCALE MAP SHOWING POSITIONS OF PEOPLE AND ITEMS ON THE VARIABLE OBSERVATION FORM

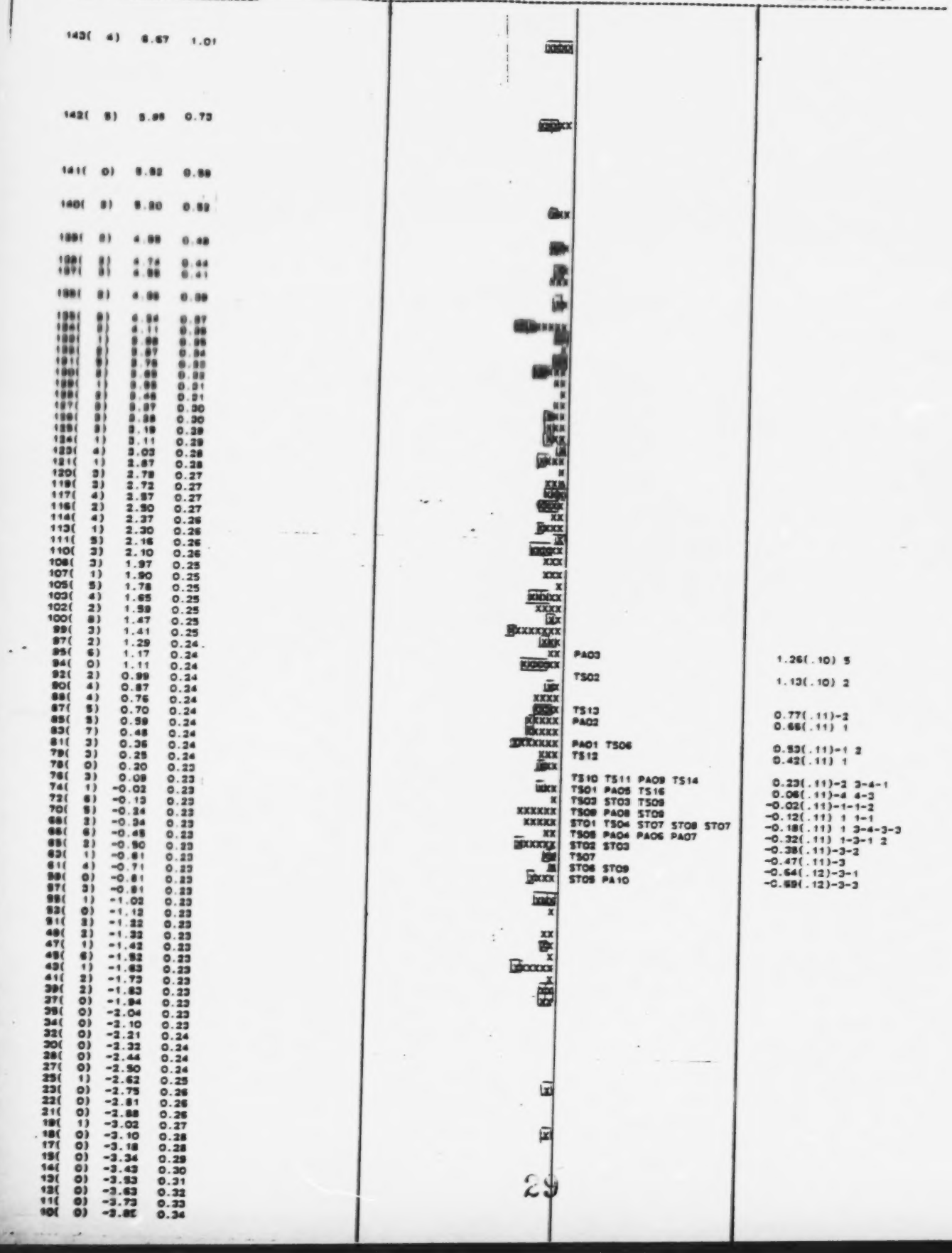


FIGURE 5

SCALE MAP SHOWING POSITIONS OF PEOPLE AND ITEMS ON THE VARIABLE
SELF-REPORT FORM

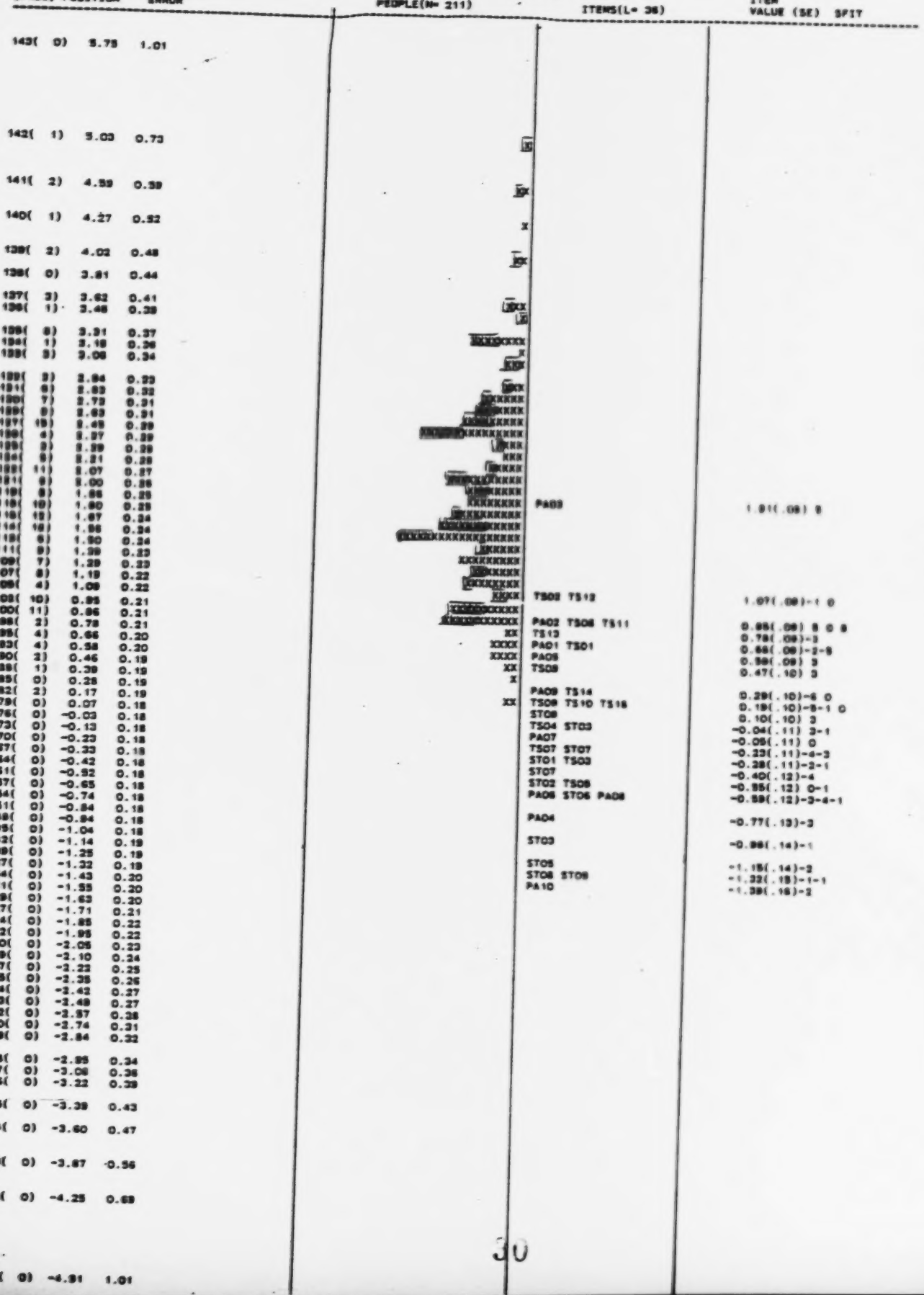
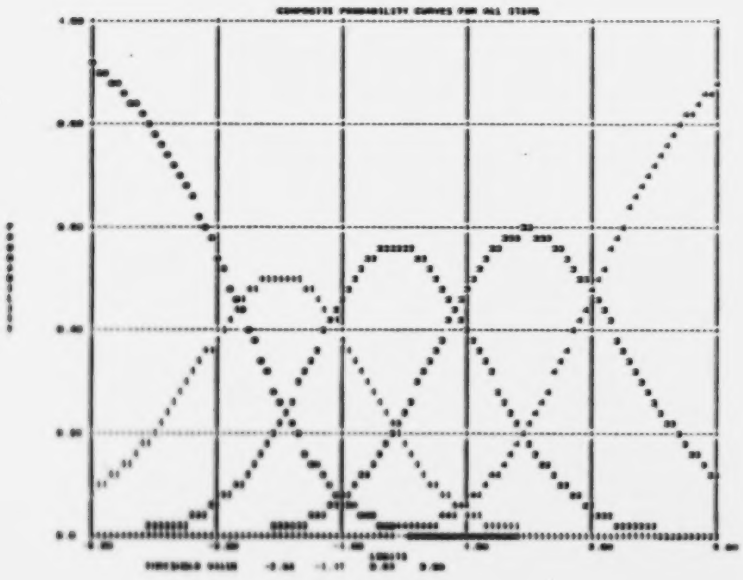


FIGURE 6

CATEGORY PROBABILITY CURVES FOR THE IMP

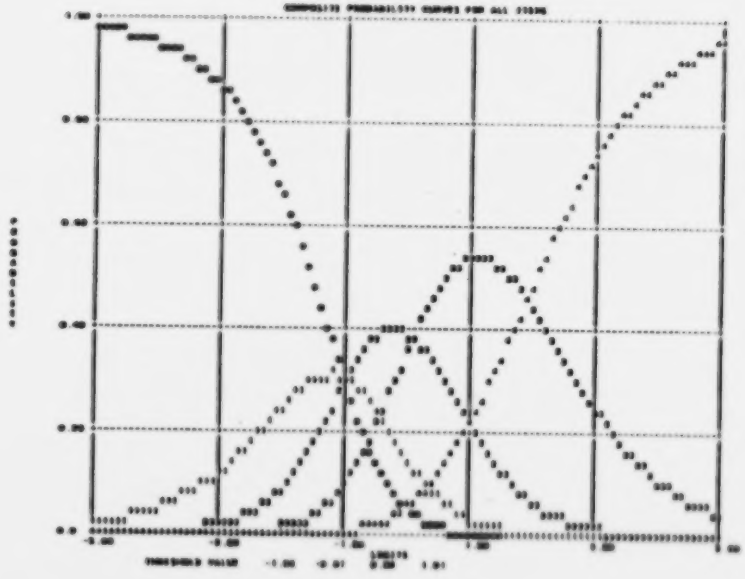
OBSERVATION FORM



204 People (M = 1.72, S = 2.07) 36 Items

Max. of 5 Categories 144 Steps

SELF-REPORT FORM



211 People (M = 1.93, S = 0.87) 36 Items

Max. of 5 Categories 144 Steps